

Adaptive Shortest-Path Routing under Unknown and Stochastically Varying Link States

Keqin Liu, Qing Zhao

Dept. of Elec. and Comp. Eng., Univ. of California, Davis

Email: {kqliu,qzhao}@ucdavis.edu

Abstract

We consider the adaptive shortest-path routing problem in wireless networks under unknown and stochastically varying link states. In this problem, we aim to optimize the quality of communication between a source and a destination through adaptive path selection. Due to the randomness and uncertainties in the network dynamics, the quality of each link varies over time according to a stochastic process with unknown distributions. After a path is selected for communication, the aggregated quality of all links on this path (*e.g.*, total path delay) is observed. The quality of each individual link is not observable. We formulate this problem as a multi-armed bandit with dependent arms. We show that by exploiting arm dependencies, a regret polynomial with network size can be achieved while maintaining the optimal logarithmic order with time. This is in sharp contrast with the exponential regret order with network size offered by a direct application of the classic MAB policies that ignore arm dependencies. Furthermore, our results are obtained under a general model of link-quality distributions (including heavy-tailed distributions) and find applications in cognitive radio and ad hoc networks with unknown and dynamic communication environments.

I. INTRODUCTION

We consider a wireless network where the quality state of each link is unknown and stochastically varying over time. Specifically, the state of each link is modeled as a random cost (or reward) evolving i.i.d. over time under an unknown distribution. At each time, a path from the source to the destination is selected as the communication route and the *total end-to-end cost*, given by the sum of the costs of all links on the path, is subsequently observed. The cost of each individual link is not observable. The objective is to design the optimal sequential path selection policy to minimize the long-term total cost.

A. Stochastic Online Learning based on Multi-Armed Bandit

The above problem can be modeled as a generalized Multi-Armed Bandit (MAB) with dependent arms. In the classic MAB [1]–[6], there are N independent arms and a player needs to decide which arm to play at each time. An arm, when played, offers a random cost drawn from an unknown distribution. The performance of a sequential arm selection policy is measured by regret, defined as the difference in expected total cost with respect to the optimal policy in the ideal scenario of known cost models where the player always plays the best arm. The optimal regret was shown to be logarithmic with time in [2]. Furthermore, since cost observations from one arm do not provide information about the cost of other arms, the optimal regret grows linearly with the number of arms. We can model the adaptive routing problem as an MAB by treating each path as an arm. The difference is that paths are dependent through shared links. While the dependency across paths can be ignored in learning and the policies for the classic MAB directly apply, such a naive approach yields poor performance with a regret growing linearly with the number of paths, thus exponentially with the network size (*i.e.*, the number of links in the worst case).

In this paper, we show that by exploiting the structure of the path dependencies, a regret polynomial with network size can be achieved while preserving its optimal logarithmic order with time. Specifically, we propose an algorithm that achieves an $O(d^3 \log T)$ regret for all light-tailed cost distributions, where $d \leq m$ is the dimension of the path set, m the number of links, and T the length of time horizon. We further show that a modification to the proposed algorithm leads to a regret linear with d by sacrificing an arbitrarily small regret order with time. This result allows a performance tradeoff in terms of the network size and the time horizon. For example, the algorithm with a smaller regret order in d can perform better over a finite time horizon when the network size is large. For heavy-tailed cost distributions, a regret linear with the number of edges and sublinear with time can be achieved. We point out that any regret sublinear with time implies the convergence of the time-average cost to the minimum one of the best path.

We further generalize the adaptive routing problem to stochastic online linear optimization problems, where the action space is a compact subset of \mathcal{R}^d and the random cost function is linear on the actions. We show that for all light-tailed cost functions, regret polynomial or linear with d and sublinear with time can be achieved. We point out that for cases where there exists a nonzero gap in the expected cost between the optimal action and the rest of the action space (*e.g.*, when the action set is a polytope or finite), the same regret orders obtained for the adaptive routing problem can be achieved.

B. Applications

One application example is adaptive routing in cognitive radio networks where secondary users communicate by exploiting channels temporarily unoccupied by primary users. In this case, the availability of each link dynamically varies according to the communication activities of nearby primary users. The delay on each link can thus be modeled as a stochastic process unknown to the secondary users. The objective is to route through the path with the smallest latency (*i.e.*, the lightest primary traffic) through stochastic online learning.

Other applications include ad hoc networks where link states vary stochastically due to channel fading or random contentions with other users.

C. Related Work

The classic MAB was addressed by Lai and Robbins in 1985, where they showed that the minimum regret has a logarithmic order with time and proposed specific policies to asymptotically achieve the optimal regret for several cost distributions in the light-tailed family [2]. Since Lai and Robbins's seminar work, simpler index policies were proposed for different classes of light-tailed cost distributions to achieve the logarithmic regret order with time [3]–[5]. In [6], we proposed the DSEE approach that achieves the logarithmic regret order with time for all light-tailed cost distributions and sublinear regrets with time for heavy-tailed cost distributions. All regrets established in [2]–[6] grow linearly with the number of arms. This work builds upon the general DSEE structure proposed in [6]. By incorporating the exploitation of the arm dependencies into DSEE, we show that the learning efficiency can be significantly improved in terms of network size while preserved in terms of time.

This work generalizes the previous work [7]–[9] that assumes fully observable link costs on a chosen path. In particular, the adaptive routing problem under this more informative observation model was considered in [9] and an algorithm was proposed to achieve $O(m^4 \log T)$ regret, where m is the number of links in the network. The problems considered in this paper were also studied under an adversarial bandit model in which the cost functions are chosen by an adversary and are treated as arbitrary bounded deterministic quantities [10]. Algorithms were proposed to achieve regrets sublinear with time and polynomial with network size. The problem formulation and results established in this paper can be considered as a stochastic version of those in [10].

For the special class of stochastic online linear optimization problems where there exists a nonzero gap in expected cost between the optimal action and the rest of the action space and the cost has a finite support, an algorithm was proposed in [11] to achieve an $O(d^2 \log^3 T)$ regret given that a nontrivial

lower bound on the gap is known. The algorithm proposed in this paper achieves a better regret in terms of both d and T without any knowledge on the cost model. For the general case with finite-support cost distributions, the regret was shown to be lower bounded by $O(d\sqrt{T})$ and an efficient algorithm was proposed to achieve an $O((d\ln T)^{3/2}\sqrt{T})$ regret [11]. Compared to the algorithm in [11], our algorithm for the general stochastic online linear optimization problems performs worse in T but better in d .

II. PROBLEM STATEMENT

In this section, we address the adaptive routing problem under unknown and stochastically time-varying link states. Consider a network with a source s and a destination r . Let $G = (V, E)$ denote the directed graph consisting of all simple paths from s to r . Let m and n denote, respectively, the number of edges and vertices in graph G .

At each time t , a random weight/cost $W_e(t)$ drawn from an unknown distribution is assigned to each edge in E . We assume that $\{W_e(t)\}$ are i.i.d. over time for each edge e . At the beginning of each time slot t , a path $p_n \in \mathcal{P}$ ($i \in \{1, 2, \dots, |\mathcal{P}|\}$) from s to r is chosen, where \mathcal{P} is the set of all paths from s to r in G . Subsequently, the *total end-to-end cost* $C_t(p_n)$ of the path p_n , given by the sum of the weights of all edges on the path, is revealed in the end of the slot. We point out that the individual cost on each edge $e \in E$ is unobservable.

The regret of a path selection policy π is defined as the expected extra cost incurred over time T compared to the optimal policy that always selects the best path (*i.e.*, the path with the minimum expected total end-to-end cost). The objective is to design a path selection policy π to minimize the growth rate of the regret. Let σ be a permutation on all paths such that

$$\mathbb{E}[C_t(\sigma(1))] \leq \mathbb{E}[C_t(\sigma(1))] \leq \mathbb{E}[C_t(\sigma(|\mathcal{P}|))].$$

We define regret

$$\mathcal{R}^\pi(T) \triangleq \mathbb{E}\left[\sum_{t=1}^T (C_t(\pi) - C_t(\sigma(1)))\right],$$

where $C_t(\pi)$ denotes the total end-to-end cost of the selected path under policy π at time t .

III. ADAPTIVE SHORTEST PATH ROUTING ALGORITHM

In this section, we present the proposed adaptive shortest-path routing algorithm. We consider both light-tailed and heavy-tailed cost distributions.

A. Light-Tailed Cost Distributions

We consider the light-tailed cost distributions as defined below.

Definition 1: A random variable X is light-tailed if its moment-generating function exists, *i.e.*, there exists a $u_0 > 0$ such that

$$M(u) \triangleq \mathbb{E}[\exp(uX)] < \infty \quad \forall u \leq |u_0|;$$

otherwise X is heavy-tailed.

For a zero-mean light-tailed random variable X , we have [12],

$$M(u) \leq \exp(\zeta u^2/2), \quad \forall u \leq |u_0|, \quad \zeta \geq \sup\{M^{(2)}(u), \quad -u_0 \leq u \leq u_0\}, \quad (1)$$

where $M^{(2)}(\cdot)$ denotes the second derivative of $M(\cdot)$ and u_0 the parameter specified in Definition 1. From (1), we have the following extended Chernoff-Hoeffding bound on the deviation of the sample mean from the true mean for light-tailed random variables [6].

Let $\{X(t)\}_{t=1}^\infty$ be i.i.d. light-tailed random variables. Let $\overline{X}_s = (\sum_{t=1}^s X(t))/s$ and $\theta = \mathbb{E}[X(1)]$. We have, for all $\delta \in [0, \zeta u_0]$, $a \in (0, 1/(2\zeta)]$,

$$\Pr(|\overline{X}_s - \theta| \geq \delta) \leq 2 \exp(-a\delta^2 s). \quad (2)$$

Note that the path cost is the sum of the costs of all edges on the path. Since the number of edges in a path is upper bounded by m , the bound on the moment generating function in (1) holds on the path cost by replacing ζ by $m\zeta$, and so does the Chernoff-Hoeffding bound in (2).

In the following, we propose an algorithm to achieve a regret polynomial with m and logarithmic with T . We first represent each path p_n as a vector \vec{p}_n with m entries consisting of 0s and 1s representing whether or not an edge is on the path. The vector space of all paths is embedded in a d -dimensional ($d \leq m$) subspace¹ of \mathcal{R}^m . The cost on each path p_n at time t is thus given by the linear function

$$[W_1(t), W_2(t), \dots, W_m(t)] \cdot \vec{p}_n.$$

The general structure of the algorithm follows the DSEE framework established in [6] for the classic MAB. More specifically, we partition time into an exploration sequence and an exploitation sequence. In the exploration sequence, we sample the d basis vectors (barycentric spanner [10] as defined in the following) evenly to estimate the qualities of these vectors. In the exploitation sequence, we select the action estimated as the best by linearly interpolating the estimated qualities of the basis vectors. A detailed algorithm is given in Fig. 1.

¹If graph G is acyclic, then $d = m - n + 2$.

Definition 2: A set $B = \{x_1, \dots, x_d\}$ is a *barycentric spanner* for a d -dimensional set A if every $x \in A$ can be expressed as a linear combination of elements of B using coefficients in $[-1, 1]$.

Note that a compact subset of \mathcal{R}^d always has a barycentric spanner, which can be constructed efficiently (see an algorithm in [10]). For the problem at hand, the path set \mathcal{P} is a compact subset of \mathcal{R}^d with dimension $d \leq m$. We can thus construct a barycentric spanner consisting of d paths in \mathcal{P} by using the algorithm in [10].

Adaptive Shortest-Path Routing Algorithm

- Notations and Inputs: Construct a barycentric spanner $B = \{\vec{p}_1, \vec{p}_2, \dots, \vec{p}_d\}$ of the vector space of all paths. Let $\mathcal{A}(t)$ denote the set of time indices that belong to the exploration sequence up to (and including) time t and $\mathcal{A}(1) = \{1\}$. Let $|\mathcal{A}(t)|$ denote the cardinality of $\mathcal{A}(t)$. Let $\bar{\theta}_{p_i}(t)$ denote the sample mean of path p_i ($i \in \{1, \dots, d\}$) computed from the past cost observations on the path. For two positive integers k and l , define $k \oslash l \triangleq ((k-1) \bmod l) + 1$, which is an integer taking values from $1, 2, \dots, l$.
- At time t ,
 1. if $t \in \mathcal{A}(t)$, choose path p_n with $n = |\mathcal{A}(t)| \oslash d$;
 2. if $t \notin \mathcal{A}(t)$, estimate the sample mean of each path $p_n \in \mathcal{P}$ by linearly interpolating $\{\bar{\theta}_{p_1}(t), \bar{\theta}_{p_2}(t), \dots, \bar{\theta}_{p_d}(t)\}$. Specifically, let $\{a_i : |a_i| \leq 1\}_{i=1}^d$ be the coefficients such that $\vec{p}_n = \sum_{i=1}^d a_i \vec{p}_i$, then $\bar{\theta}_{p_n}(t) = \sum_{i=1}^d a_i \bar{\theta}_i(t)$. Choose path

$$p^* = \arg \min \{\bar{\theta}_{p_n}(t), n = 1, 2, \dots, |\mathcal{P}|\}.$$

Fig. 1. The general structure of the algorithm.

Theorem 1: Construct an exploration sequence as follows. Let a, ζ, u_0 be the constants such that (2) holds on each edge cost. Choose a constant $b > 2m/a$, a constant

$$c \in (0, \min_{j: \mathbb{E}[C_t(\sigma(j)) - C_t(\sigma(1))] > 0} \{\mathbb{E}[C_t(\sigma(j)) - C_t(\sigma(1))]\}),$$

and a constant $w \geq \max\{b/(md\zeta u_0)^2, 4b/c^2\}$. For each $t > 1$, if $|\mathcal{A}(t-1)| < d\lceil d^2 w \log t \rceil$, then include t in $\mathcal{A}(t)$. Under this exploration sequence, the resulting policy π^* has regret

$$R^{\pi^*}(T) \leq Amd^3 \log T$$

for some constant A independent of d, m and $T > 1$.

Proof: Since $\mathcal{A}(T) \leq d[d^2 w \log T]$, the regret caused in the exploration sequence is at the order of $md^3 \log T$. Now we consider the regret caused in the exploitation sequence. Let E_k denote the k th exploitation period which is the k th contiguous segment in the exploitation sequence. Let E_k denote the k th exploitation period. Similar to the proof of Theorem 3.1 in [6], we have

$$|E_k| \leq ht_k \quad (3)$$

for some constant h independent of d and m . Let $t_k > 1$ denote the starting time of the k th exploitation period. Next, we show that by applying the Chernoff-Hoeffding bound in (2) on the path cost, for any t in the k th exploitation period and $i = 1, \dots, d$, we have

$$\Pr(|\mathbb{E}[C_t(p_i)] - \bar{\theta}_{p_i}(t)| \geq c/(2d)) \leq 2t_k^{-ab/m}.$$

To show this, we define the parameter $\epsilon_i(t) \triangleq \sqrt{b \log t / \tau_i(t)}$, where $\tau_i(t)$ is the number of times that path i has been sampled up to time t . From the definition of parameter b , we have

$$\epsilon_i(t) \leq \min\{m\zeta u_0, c/(2d)\}. \quad (4)$$

Applying the Chernoff-Hoeffding bound, we arrive at

$$\Pr(|\mathbb{E}[C_t(p_i)] - \bar{\theta}_{p_i}(t)| \geq c/(2d)) \leq \Pr(|\mathbb{E}[C_t(p_i)] - \bar{\theta}_{p_i}(t)| \geq \epsilon_i(t)) \leq 2t_k^{-ab/m}.$$

In the exploitation sequence, the expected times that at least one path in B has a sample mean deviating from its true mean cost by $c/(2d)$ is thus bounded by

$$\sum_{k=1}^{\infty} 2dt_k^{-ab/m} t_k \leq \sum_{t=1}^{\infty} 2dt^{1-ab/m} = gd. \quad (5)$$

for some constant g independent of d and m . Based on the property of the barycentric spanner, the best path would not be selected in the exploitation sequence only if one of the basis vector in B has a sample mean deviating from its true mean cost by at least $c/(2d)$. We thus proved the theorem. ■

In Theorem 1, we need a lower bound (parameter c) on the difference in the cost means of the best and the second best paths. We also need to know the bounds on parameters ζ and u_0 such that the Chernoff-Hoeffding bound (2) holds. These bounds are required in defining w that specifies the minimum leading constant of the logarithmic cardinality of the exploration sequence necessary for identifying the best path. Similar to [6], we can show that without any knowledge of the cost models, increasing the cardinality of the exploration sequence of π^* by an arbitrarily small amount leads to a regret linear with d and arbitrarily close to the logarithmic order with time.

Theorem 2: Let $f(t)$ be any positive increasing sequence with $f(t) \rightarrow \infty$ as $t \rightarrow \infty$. Revise policy π^* in Theorem 1 as follows: include t ($t > 1$) in $\mathcal{A}(t)$ if $|\mathcal{A}(t-1)| < d\lceil f(t) \log t \rceil$. Under the revised policy π' , we have

$$R^{\pi'}(T) = O(df(T) \log T).$$

Proof: It is sufficient to show that the regret caused in the exploitation sequence is bounded by $O(d)$, independent of T . Since the exploration sequence is denser than the logarithmic order as in Theorem 1, it is not difficult to show that the bound on $|E_k|$ given in (3) still holds with a different value of h .

We consider any positive increasing sequence $b(t)$ such that $b(t) = o(f(t))$ and $b(t) \rightarrow \infty$ as $t \rightarrow \infty$. By replacing b in the proof of Theorem 1 with $b(t)$, we notice that after some finite time T_0 , the parameter $\epsilon_i(t)$ will be small enough to ensure (4) holds and $b(t)$ will be large enough to ensure (5) holds. The proof thus follows. \blacksquare

B. Heavy-Tailed Cost Distributions

We now consider the heavy-tailed cost distributions where the moment of the cost exists up to the q -th ($q > 1$) order. From [6], we have the following bound on the deviation of the sample mean from the true mean for heavy-tailed cost distributions.

Let $\{X(t)\}_{t=1}^\infty$ be i.i.d. random variables drawn from a distribution with the q -th moment ($q > 1$). Let $\bar{X}_t = (\sum_{k=1}^t X(k))/t$ and $\theta = \mathbb{E}[X(1)]$. We have, for all $\epsilon > 0$,

$$\Pr(|\bar{X}_t - \theta| > \epsilon) = o(t^{1-q}). \quad (6)$$

Theorem 3: Construct an exploration sequence as follows. Choose a constant $v > 0$. For each $t > 1$, if $|\mathcal{A}(t-1)| < vt^{1/q}$, then include t in $\mathcal{A}(t)$. Under this exploration sequence, the resulting policy π^q has regret $R^{\pi^q}(T) \leq DdT^{1/q}$ for some constant D independent of d and T .

Proof: Based on the construction of the exploration sequence, it is sufficient to show that the regret in the exploitation sequence is $o(T^{1/q}) \cdot d$. From (6), we have, for any $i = 1, \dots, d$,

$$\Pr(|\mathbb{E}[C_t(p_i)] - \bar{\theta}_{p_i}(t)| \geq c/(2d)) = o(|\mathcal{A}(t)|^{1-q}).$$

For any exploitation slot $t \in \mathcal{A}(t)$, we have $|\mathcal{A}(t)| \geq vt^{1/q}$. We arrive at

$$\Pr(|\mathbb{E}[C_t(p_i)] - \bar{\theta}_{p_i}(t)| \geq c/(2d)) = o(t^{(1-q)/q}).$$

Since the best path will not be chosen only if at least one of the basis vector has the sample deviating from the true mean by $c/(2d)$, the regret in the exploitation sequence is thus bounded by

$$\sum_{t=1}^T o(t^{(1-q)/q}) \cdot d = o(T^{1/q}) \cdot d.$$

IV. GENERALIZATION TO STOCHASTIC ONLINE LINEAR OPTIMIZATION PROBLEMS

In this section, we consider the general Stochastic Online Linear Optimization (SOLO) problems that include adaptive routing as a special case. In a SOLO problem, there is a compact set $\mathcal{U} \subset \mathcal{R}^d$ with dimension d , referred to as the action set. At each time t , we choose a point $x \in \mathcal{U}$ and observe a cost $\vec{C}_t \cdot x$ where $\vec{C}_t \in \mathcal{R}^d$ is drawn from an unknown distribution². We assume that $\{\vec{C}_t\}_{t \geq 1}$ are i.i.d. over t . The objective is to design a sequential action selection policy π to minimize the regret $\mathcal{R}^\pi(T)$, defined as the total difference in the expected cost over T slots compared to the optimal action $x^* \triangleq \arg \min\{\mathbb{E}[\vec{C}_t \cdot x]\}$:

$$\mathcal{R}^\pi(T) \triangleq \mathbb{E}\left[\sum_{t=1}^T (\vec{C}_t \cdot x(\pi) - \vec{C}_t \cdot x^*)\right],$$

where $x(\pi)$ denotes the action under policy π at time t .

As a special case, the action space of the adaptive routing problem consists of all path vectors with dimension $d \leq m$. The main difficulty in a general SOLO problem is on identifying the optimal action in an infinite action set. Our basic approach is to implement the shortest-path algorithm in Fig. 1 repeatedly for the chosen action to gradually converge to the optimal action. Specifically, the proposed algorithm runs under an epoch structure where each epoch simulates one round of the algorithm and the epoch lengths form a geometric progression.

Theorem 4: Assume that the cost for each action is light-tailed. Let T_k be the length of the k -th epoch. Construct an exploration sequence in this epoch as follows. Let a, ζ, u_0 be the constants such that (2) holds on each cost. Choose a constant $b > 2/a$, a constant $c = \log T_k^{1/3}/T_k^{1/3}$, and a constant $w \geq \max\{b/(d\zeta u_0)^2, 4b/c^2\}$. For each $t > 1$, if $|\mathcal{A}(t-1)| < d\lceil d^2 w \log t \rceil$, then include t in $\mathcal{A}(t)$. The resulting policy π^g has regret

$$R^{\pi^g}(T) = O(d^3 T^{2/3} \log^{1/3} T).$$

We point out that the regret order can be improved to linear with d by sacrificing the order with T by an arbitrarily small amount, as similar to Theorem 2.

V. CONCLUSION

In this paper, we considered the adaptive routing problem in networks with unknown and stochastically varying link states, where only the total end-to-end cost of a path is observable after the path is selected

²The following result holds in a more general scenario that only requires the expected cost is linear with x .

for routing. For both light-tailed and heavy-tailed link-state distributions, we proposed efficient online learning algorithms to minimize the regret in terms of both time and network size. The result was further extended to the general stochastic online linear optimization problems. Future work includes extending the i.i.d. cost evolution over time to more general stochastic processes.

REFERENCES

- [1] H. Robbins, "Some Aspects of the Sequential Design of Experiments," *Bull. Amer. Math. Soc.*, vol. 58, no. 5, pp. 527-535, 1952.
- [2] T. Lai and H. Robbins, "Asymptotically Efficient Adaptive Allocation Rules," *Advances in Applied Mathematics*, vol. 6, no. 1, pp. 4-22, 1985.
- [3] T. Lai, "Adaptive Treatment Allocation and The Multi-Armed Bandit Problem," *Ann. Statist.*, vol 15, pp. 1091-1114, 1987.
- [4] R. Agrawal, "Sample Mean Based Index Policies with $O(\log n)$ Regret for the Multi-armed Bandit Problem," *Advances in Applied Probability*, vol. 27, pp. 1054-1078, 1995.
- [5] P. Auer, N. Cesa-Bianchi, P. Fischer, "Finite-time Analysis of the Multiarmed Bandit Problem," *Machine Learning*, vol. 47, pp. 235-256, 2002.
- [6] K. Liu and Q. Zhao, "Multi-Armed Bandit Problems with Heavy Tail Reward Distributions," in *Proc. of Allerton Conference on Communications, Control, and Computing*, September, 2011.
- [7] A. Kalai and S. Vempala, "Efficient Algorithms for Online Decision Problems," *J. Computer and System Sciences*, vol. 71, pp. 291-307, 2005.
- [8] E. Takimoto, M. Warmuth, "Path Kernels and Multiplicative Updates," *Journal of Machine Learning Research*, vol. 4, pp. 773-818, 2003.
- [9] Y. Gai, B. Krishnamachari, and R. Jain, "Combinatorial Network Optimization with Unknown Variables: Multi-Armed Bandits with Linear Rewards," *USC Technical Report*, CENG-2010-9.
- [10] B. Awerbuch, R. Kleinberg, "Online Linear Optimization and Adaptive Routing," *Journal of Computer and System Sciences*, pp. 97-114, 2008.
- [11] V. Dani, T. Hayes, S. Kakade, "Stochastic Linear Optimization under Bandit Feedback," in *Proc. of the 21st Annual Conference on Learning Theory*, 2008.
- [12] P. Chareka, O. Chareka, S. Kennedy, "Locally Sub-Gaussian Random Variable and the Strong Law of Large Numbers," *Atlantic Electronic Journal of Mathematics*, vol. 1, no. 1, pp. 75-81, 2006.